https://doi.org/10.31489/2025FEB2/62-74

UDC 633.51 Received: 9 February 2025 | Accepted: 28 March 2025

A. Orken[1], Sh. Manabayeva[2], S. Makhmadjanov[3], M. Ramazanova[4], B. Kali[5], L. Tokhetova[6], D. Tussipkan[7*]

[1,2,4,5,7]*National Center for Biotechnology, Korgalzhyn hwy. 13/5, Astana, 010000, Kazakhstan;*
[3,6]*Agricultural Experimental Station of Cotton and Melon Growing, Atakent, 160525 Kazakhstan;*
[1,2]*L.N. Gumilyov Eurasian National University, Satpayev st. 2, Astana, 010000, Kazakhstan*
[*]*Corresponding author: Tussipkan Dilnur tdilnur@mail.ru*

Cotton (*Gossypium* L.) production and importance of sequencing technology for improving agronomic traits

Gossypium L. is one of the largest genera, known for its diversity and economic value among field crops, while allotetraploid cotton species have a valuable source of a model system for studying plant polyploidy, phylogeny, and breeding. This review provides, first, the production and use of Cotton (*Gossypium*) in the world. Second, important information on cotton cultivation and production in Kazakhstan was provided in detail. Third, we summarized the phylogeny of *Gossypium* L. Fourth, we provided a brief summary of morphological characteristics and whole genome sequence studies of seven allotetraploid cotton species including *G. hirsutum* (AD)1, *G. barbadense* (AD)2, *G. tomentosum* (AD)3, *G. mustelinum* (AD)4, *G. darwinii* (AD)5, *G. ekmanianum* (AD)6 and *G. stephensii* (AD)7. This review is valuable for future agronomic and molecular research studies on cotton.

Keywords: Cotton (*Gossypium* L.) production, phylogenesis, allotetraploid species, whole genomic sequences.

1 Production and uses of cotton (Gossypium L.)

Cotton (Gossypium L.) has been cultivated for fiber production more than 7000 years. Despite the presence of synthetic fibers derived from petroleum, it continues to serve as the most important natural renewable source in the world for textiles. Cotton is mainly grown in more than 80 countries around the world, including China, India, USA and Pakistan [1, 2].

Furthermore, cotton is the main economic driver for some developing countries. In addition to fiber, cotton is the third largest arable crop in the world in terms of tons of edible oilseed after soybeans and rapeseed. In addition to its 21 % fat content, cottonseed is a source of relatively high quality protein.

Industrial cotton species include diploids (*G. herbaceum* and *G. arboreum*) and tetraploids (*G. barbadense* and *G. hirsutum*). The origin of both diploid species is considered to be South Asia and Africa, while the origin of two allotetraploid species are considered to be from Central, North, and South America. Among these four cultivated species, *G. hirsutum* has high yield potential, wide adaptability, and moderate fiber quality and accounts for about 90–95 % of the total cotton production [3].

The cotton industry is an important part of the economy of several countries. In India, the cotton sector employs more than 40 million people, including farmers, workers from refining and pressing plants, and workers from textile factories. The US cotton industry provides about 250,000 jobs and contributes about 21 billion to the economy each year. Cotton farming also plays an important role in the socio-economic structure of countries such as Pakistan and Brazil, providing a livelihood for millions of families [1].

Cotton is grown mainly in Asia (~70 %), followed by the Americas (20 %), Africa (6 %), Europe (2 %), and other regions (~1 %) (Fig. 1).

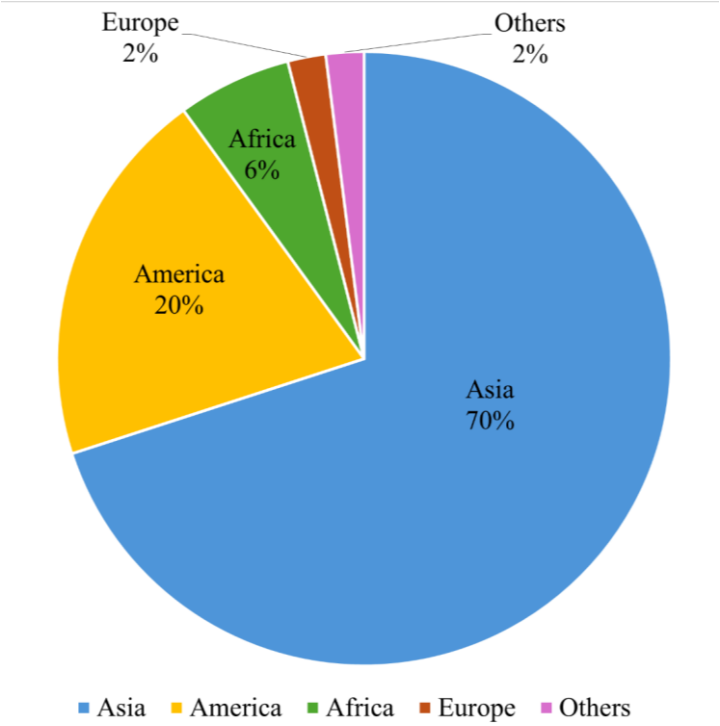

Figure 1. Trend in cotton cultivation around the world

China, India, the United States, Brazil, Australia, Turkey, Pakistan, Uzbekistan, Argentina, and Mali are known as world's leading cotton producing countries in 2022 and 2023 (in 1,000 metric tons) (Fig. 2).

According to foreign media reports, cotton production in 2023 will reach 24.123 million tons, an increase of 6.2 % over previous years, according to data published by the International Cotton Advisory Committee (ICAC). The projected volume of world cotton production for 2024-2025 is expected to reach 25.6204 million tons. World cotton production for the 2023-24 season is projected to increase by 3 % to 25.41 million tons, while consumption is expected to decrease by 0.43 % to 23.35 million tons [5]

Seed cotton goes through a disassembly process in ginning factories to produce fiber, cottonseed, and waste. Cotton fiber is a raw material for the textile industry. Oil can be extracted from the seeds (about 20 % of the seed mass is fat), it can be a vegetable oil, margarine, soap, etc. Several by-products of seed cotton are obtained along with the main product — fiber. After ginning, the by-products are used as animal feed and for the production of biofuels [6]. Cotton consumption in China will decrease by 15 %, and India will become the largest consumer with 6.7 million tons. Consumption growth is also expected to increase in Vietnam, Bangladesh, Indonesia, and Turkey [7].

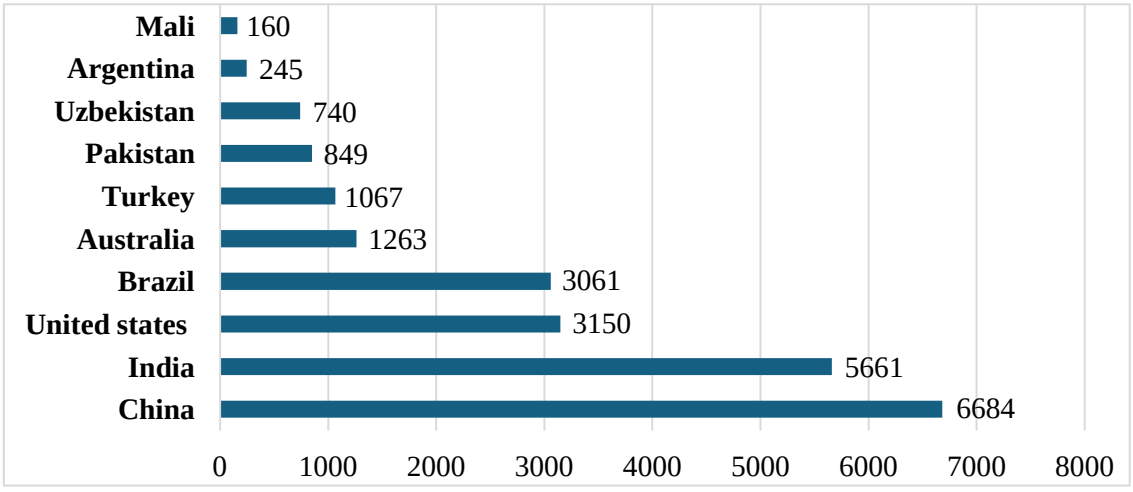

Figure 2. Leading cotton producing countries of the world in 2022 and 2023 years [4]

1.1 Cotton production in Kazakhstan

The Turkestan region is the world's northernmost cotton-growing region in southern Kazakhstan. Every year, 115,000–125,000 hectares of medium-staple cotton (G. hirsutum L.) are planted, with 80,000–85,000 hectares being in Maktaaral and Zhetysay Districts of Kazakhstan. This area is particularly vulnerable to drought, salt, and the invasion of harmful pests, including beetroot borer, cotton budworm, spider mites, and aphids, as well as illnesses like gummosis and fusarium blight (wilt). Through genotype selection based on genetic principles, the adverse effects of extremely high salinity and aridity in arable soil can be effectively and economically mitigated [8]. The dynamics of sown areas in the Turkestan region have shown significant changes over the past decades (Fig. 3). In order to better understand the dynamics of sown areas in the Turkestan region, we have divided the analysis into five distinct periods: 1991–1997, 1998–2004, 2005–2010, 2011–2015, and 2016–2023 years.

In 1991–1997 years, according to the analysis of sown area, the sown area decreased from 116.6 thousand hectares to 103.6 thousand hectares, with an average of 112 thousand hectares.

In 1998–2004 years, the sown area was increased significantly from 118 thousand hectares to 223.1 thousand hectares, with an average of 118 thousand hectares.

In 2005–2010 years, the sown area significantly decreased from 206.1 thousand hectares to 137.3 thousand hectares, with an average of 172 thousand hectares.

In 2011–2015 years, the sown area did not show stable characteristics and decreased from 160.6 thousand hectares to 99.3 thousand hectares, with an average of 135,18 thousand hectares.

In 2016–2023 years, the sown area showed increasing characteristics from 109,6 thousand hectares to 135,5 thousand hectares, with an average of 123,4 thousand hectares.

Figure 3. 1991–2023 years, dynamics of sowing areas in the Turkestan region of Kazakhstan (thousand hectares)

Out of the 13 cotton varieties developed by the "Agricultural Experimental Station of Cotton and Melon Growing" LLP, 9 varieties — Bereke-07, Maktaaral-4005, Maktaaral-4007, Maktaaral-4011, Myrzashol-80, Pakhtaaral-3031, Pakhtaaral-3044, Maktaaral-4017, and Maktaaral-5027 — have been included in the Register of Breeding Achievements recommended for use in the Republic of Kazakhstan. These varieties have been widely adopted and cover more than 90 % of the cotton growing areas in the Turkestan region. All 13 varieties have received patents for breeding achievements. In 2021, a patent was granted for a new promising cotton variety, Maktaaral-5027, which has relatively high resistance to pests such as cotton bollworm and spodoptera [8].

 ISSN 3080-6836 (Print) ISSN 3080-6844 (Online)

The major challenges in cotton production include salinization, irrigation water scarcity, climate change, pests and diseases, and soil depletion. In response, breeding efforts are focused on developing crops with early maturity (within 105–110 days), high salt tolerance, drought resistance, diseases and pest resistance, and the ability to improve soil fertility in crop rotation systems.

Currently, the Agricultural Experimental Station of Cotton and Melon Growing maintains a cotton gene pool of 700 samples from different countries around the world as of 2024.

The soil of the experimental area is light fertile gray soil with a medium loam mechanical composition. The characteristics of light fertile grey soil include a low humus content, high carbonate content, and a relatively low absorptive capacity.

This soil type is characterized by good microstructure, water permeability, porosity, relatively low cohesiveness, and moderate mobility of water and nutrients.

2 Phylogenesis and whole genome sequences of cotton

The genus cotton (*Gossypium L.*) not only has the highest economic value among field crops, but also plays an important role in research on plant taxonomy, polyploidization, phylogeny, cytogenetics, and genomics.

Gossypium L. belongs to the tribe *Gossypieae*, the family *Malvaceae*, and the order *Malvales* [9]. Species belonging to the genus *Gossypium* are grouped into 8 diploid genomes (A to G and K) and one allopolyploid genome (AD) based on relative chromosome sizes and chromosome features in interspecific hybrids [10]. A fairly well-documented review article has been published on the taxonomic, cytogenetic, and geographic diversity of *Gossypium* species and the nomenclature of individual genomes and chromosomes [11].

2.1 Genomes of allotetraploid species

The new allotetraploid or amphidiploid cotton (AD genome) is thought to have emerged from an ancient fusion between the A genome of the African Species *G. arboreum* and the D genome American Species *G. raimondii*, and chromosome duplication from their ancestors. The polyploids were found to contain 7 species, namely *G. hirsutum* (AD)1, *G. barbadense* (AD)2, *G. tomentosum* (AD)3, *G. mustelinum* (AD)4, *G. darwinii* (AD)5, *G. ekmanianum* (AD)6 and *G. stephensii* (AD) 7 [11]. Almost complete genome sequences have been determined for all these species. The aim of this review to provide a comparative overview of the characteristics of allotetraploid species and the features of their genomes based on previously published studies.

2.1.1 Gossypium hirsutum

G. hirsutum has 14 different Latin synonyms: G. hirsutum var. punctatum, G. jamaicense, G. lanceolatum, G. mexicanum, G. morrillii, G. punctatum, G. purpurascens, G. religiosum, G. schottii, G. taitense, G. tridens, G. tricuspidatum, G. hirsutum var. marie-galante and G. hirsutum var. palmeri. It is colloquially referred to as American tetraploid, American upland cotton, Mexican cotton and upland cotton.

G. hirsutum is the most widely cultivated cotton species and the dominant source of natural plant fiber in the world [12]. It originated in Central America, but it is cultivated worldwide. Due to its high productivity and quality, it is an important species from the perspective of breeders and serves as the primary material for scientific research aimed at combining high yield with other valuable traits. It produces medium-length fibers and can grow to a height of approximately 2 meters. During development, G. hirsutum typically has 3 to 5 lobed leaves that are generally flat and diaheliotropic, requiring sunlight to maximize light absorption.

The benefits of genomic sequencing and resequencing data are enabling the study of the genetic basis of G. hirsutum. In 2015, for the first time, the complete genome sequence of G. hirsutum cotton was developed [13]. According to the results of this study, the allotetraploid genome of *G. hirsutum* TM-1 was estimated to be 2.25–2.43 Gb using various methods. This genome was found to consist of a total of 44 816 contigs, 8591 scaffolds, 76,943 annotated protein-coding genes, 602 microRNAs (miRNAs), 2153 ribosomal RNAs (rRNAs), 2050 transfer RNAs (tRNAs), and 8325 small nuclear RNAs (snRNAs). Comparative transcriptomic studies in this research revealed the crucial role of nucleotide-binding site (NBS)-encoding genes in *Verticillium dahliae* resistance and the involvement of ethylene in cotton fiber cell development. Following this study, the complete genome sequence of *G. hirsutum* has been updated and refined several times [14–22]

According to the study by Chen, Sreedasyam et al. (2020), accession TM-1 had a genome size of 2.3 Gb with genome coverage of 94.06x and 75 376 genes [18]. The GC content is 34.5 %. Chromosome D09 had the smallest nucleotide of 54 445 796 bp and chromosome A06 had the largest nucleotide of 128 195 338 bp (Table).

2.1.2 Gossypium barbadense

G. barbadense has eight Latin synonyms, including G. peruvianum, G. vitifolium, G. acuminatum, G. barbadense var. acuminatum, G. barbadense var. brasiliense, G. brasiliense, G. guyanense var. braziliense, and G. evertum. It is colloquially known as Pima cotton, American Pima cotton, Sea Island cotton, long staple cotton, Egyptian cotton, and Brazilian cotton. It originated in South America, but it is cultivated worldwide. G. barbadense is a tropical cotton plant that can grow to the size of a small tree. It is highly valued for its long, high-quality fibers and its resistance to Verticillium wilt disease [23].

In 2015, the complete genome sequence of. *G barbadense* was completed. The genome of *G. barbadense* was found to be 2.57 gigabases, including a high-quality set of 80 876 protein-coding genes, along with information on the dynamic changes of genes and their expression [24].

In 2019, the complete genome sequence of *G. barbadense* was determined using the PacBio RSII method. The genome was found to contain 71 297 protein-coding genes with a contig length of 2 222 525 789 base pairs [16].

The study by Chen, Sreedasyam et al. (2020) found that the genome of *G. hirsutum* variety 3–79 has a genome size of 2.2 Gb, containing 74 561 genes with a coverage of 90.1x [18]. The GC content is 34 %. The smallest nucleotide sequence was found on chromosome D09, with 50 685 742 base pairs, while the largest nucleotide sequence was found on chromosome A08, with 119 114 718 base pairs (Table). Ma, Zhang et al. (2021). The genome of *G. barbadense* was presented with a high quality assembly of 2.57 gigabases, including the identification of 80 876 protein-coding genes [3].

2.1.3 Gossypium tomentosum

G. tomentosum has unique agronomic characteristics, including strong fibers, hairy leaves and stems, resistance to pests or insects on the leaves, and heat tolerance, which are traits of the *Gossypium* species. These excellent agronomic properties of *G. tomentosum* can be introduced into *G. hirsutum* through interspecific hybridization, allowing for its use in the genetic selection of *G. hirsutum* by expanding its genetic diversity [25]. Gossypium tomentosum, known as Ma'o, huluhulu, or Hawaiian cotton, is a cotton species native to the Hawaiian Islands. Genetic studies have shown that Hawaiian cotton belongs to the American species of the genus *Gossypium*, with its closest relative being *G. hirsutum*. Focusing on specific research, a 2016 study compared two genetic linkage maps based on F2 hybrids of *G. hirsutum x G. tomentosum and G. hirsutum x G. darwinii*. Seven inverted fragments were found on chromosomes chr02, chr05, chr08, chr12, chr14, chr16, and chr25, and three translocated fragments were identified on chr05, chr14, and chr26. These results indicate that *G. tomentosum* is genetically closer to *G. hirsutum* than to *G. darwinii* [26].

G. tomentosum is a shrub that grows to a height of 0.46 to 1.52 meters and has a diameter ranging from 1.5 to 3.0 meters. The plant's seed fibers (lints) are short and red-brown in color, making them unsuitable for spinning or turning into thread. Its flowers are light yellow with 3–5 petals and bloom from late summer to early winter. It is characterized by heat resistance, resistance to harmful beetles, flea beetles, weevil rot, and worms, as well as resistance to jassids and thrips. Additionally, it is known for its high quality fibers, fiber length, and fiber fineness [23].

The study by Chen, Sreedasyam et al. (2020) found that the genome of G. *tomentosum* varieties 7179.01, 7179.02, and 7179.03 has a genome size of 2.2 Gb, containing 78 281 genes with a coverage of 76.8x. The GC content is 34 %. The smallest nucleotide sequence was found on chromosome D09, with 51 553 955 base pairs, while the largest nucleotide sequence was found on chromosome A06, with 121 609 178 base pairs (Table).

2.1.4 Gossypium mustelinum

G. mustelinum is the only cotton species native to Brazil, and it is typical of the semi-arid regions in the northeastern part of the country [27]. It is a shrub-like plant that grows primarily in seasonally dry tropical

biomes [28]. In 2013, Brazilian scientists genotyped two hundred eighteen mature *G. mustelinum* plants using SSR markers and found high genetic diversity among the populations. The results of this study indicated that *G. mustelinum* plants have a high level of inbreeding and low observed heterozygosity, suggesting that the populations reproduce primarily through self-fertilization and cross-pollination between related individuals [27]. The genomic and genetic resources of *G. mustelinum* have proven effective in identifying genes for both qualitative and quantitative traits. *Gossypium* mustelinum represents the earliest divergent evolutionary lineage of *Gossypium* polyploids, and its cotton varieties possess a gene pool rich in many essential traits that have been lost in other cotton species [29]. For example, in 2020, an interspecific extended-fixed population of *G. mustelinum* × *G. hirsutum* was developed, and more than a one hundred QTLs for fiber quality traits were mapped [30]

The study by Chen, Sreedasyam et al. (2020) found that the genome of *G. tomentosum* has a genome size of 2.3 Gb, containing 74,660 genes with a coverage of 94.1x. The GC content is 34.5 %. The smallest nucleotide sequence was found on chromosome D09, with 53 194 044 base pairs, while the largest nucleotide sequence was found on chromosome A08, with 129 182 752 base pairs (Table).

2.1.5 Gossypium darwinii

Gossypium darwinii, or Darwin's cotton, is a species of cotton plant found only in the Galapagos Islands. It is characterized by several excellent traits, including fine fibers, drought tolerance, and resistance to *Fusarium* and *Verticillium* wilt diseases. Genetic studies indicate that it is closely related to the native American species *Gossypium barbadense*, suggesting that its seeds may have been dispersed from South America by wind, bird droppings, or marine debris [26].

According to the study by Chen, Sreedasyam et al. (2020), the sequenced isolate (AD)5-032 has a genome size of 2.2 Gb with a genome coverage of 80.6x and 78 303 genes. The GC content is 34 %. Chromosome D09 had the smallest nucleotide number of 52 096 622bp and chromosome A08 had the largest nucleotide number as 120 009 936 bp (Table).

The cultivated species *G. hirsutum* accounts for 90 % of world cotton production. However, its narrow genetic base limits the improvement of modern *G. hirsutum* cultivars. In contrast, the abundant genetic diversity found in wild species, such as *G. darwinii*, provides valuable resources to address this issue.

Here we would like to review some key studies as examples. For example, an interspecific high-density linkage map of *G. hirsutum* × *G. darwinii* was constructed by Chen, Khan et al. (2015), using an F2 population based entirely on genome-wide simple sequence repeat (SSR) markers [38]. As a result of this study, a total of 2,763 markers were mapped across 26 linkage groups (chromosomes), covering a genome length of 4,176.7 cm, with an average inter-locus distance of 1.5 cm. The map will offer essential information regarding the origin and evolution of the cotton genus, along with insights into genome structure and function. Additionally, it will aid in cotton genome assembly, fine mapping, map-based cloning, and the utilization of genetic germplasm from *G. darwinii* through marker-assisted selection. In the study by Xu, Ilyas et al. (2022), the RNA-seq transcriptome analysis revealed that a total of 32,693 up-regulated genes and 25,919 down-regulated genes were differentially expressed [37]. Gene ontology and KEGG pathway analyses revealed that the upregulated genes were associated with all gene ontology terms, as well as molecular functions, biological processes, and cellular components, which were significantly related to enhancing drought stress tolerance. In the study Wang, Li et al. (2024), a chromosome segment substitution line population of 553 individuals was created using *G. darwinii* × *G. darwinii*. As the result, three candidate genes were identified for three stable QTLs, including GH_A01G1096 (ARF5) and GH_A10G0141 (PDF2) associated with lint percentage, and GH_D01G0047 (KCS4) associated with seed index or oil content [36].

These findings enhance our understanding of the molecular regulatory mechanisms development of cotton breeding and provide valuable insights for marker-assisted genetic improvement in cotton.

2.1.6 Gossypium ekmanianum

Gossypium ekmanianum has three synonyms name such as *Gossypium hirsutum* var. *ekmanianum* (Wittm.) Roberty, *Gossypium latifolium* var. *ekmanianum* (Wittm.) Roberty, and *Gossypium tricuspidatum* var. *ekmanianum* (Wittm.) Mauer. Furthermore, there are also three common names Ekman's Cotton, Ekman's *Gossypium*, and Ekman's Wild Cotton. The native range of *G. ekmanianum* species is the southwest

Dominican Republic. It is a shrub and grows primarily in the seasonally dry tropical biome. *G. ekmanianum* has small, white flowers with five petals. The seeds are small, round, and brown [31] (Table).

According to the study by Peng, Xu et al. (2022), the sequenced isolate AD6 had a genome size of 2.3 Gb with genome coverage 106.0x and 74 178 genes. GC content is 34.5 %. Chromosome D09 had the smallest nucleotide number with 54 144 365 bp and chromosome A11 had the largest nucleotide number with 132 145 079 bp (Table) [35].

2.1.7 Gossypium stephensii

Its historical importance dates back to 1966, when Stephens (1966), an eminent natural historian, evolutionary geneticist, and cotton biologist, examined Wake Island cotton in his study of oceanic dispersal and identified it as a wild form of *G. hirsutum*. He noted that "Wake Island cotton does not closely resemble either Caribbean or other Pacific varieties. Stephens emphasized its distinctive characteristics, including its sprawling shrub-like growth habit, dense hairy pubescence, and larger-than-average petal spot when compared to other Pacific cottons [32]. Also, in 1992 years, Paul A. Fryxell provided a revised taxonomic interpretation of *Gossypium L*. and confirmed the distinguishing characteristics of the Wake Atoll forms in comparison with *G. hirsutum* [33]. In 2017, American scientists Gallagher, Grover, et al. (2017) reported that a new species of cotton from Wake Atoll was described as a new species of *Gossypium* by considering morphological distinctions, geographic isolation, and new molecular data including both nuclear and chloroplast genome sequence data. The morphological characteristics of the new species were well described and were named *Gossypium Stephens* after S.G. Stephens, the eminent natural historian, evolutionary geneticist, and cotton biologist [34].

In 2022, according to Peng, Xu et al. (2022) study, the sequenced isolate AD7 has a genome size of 2.3Gb with genome coverage of 127.0x and gene number are 74,970. The GC content is 34.5 %. Chromosome D09 had the smallest nucleotide number of 54 019 951bp and chromosome A06 had the largest nucleotide number of 125 976 056 bp (Table) [35].

 ISSN 3080-6836 (Print) ISSN 3080-6844 (Online)

Table

Analysis of taxonomic, geographic diversity, and genomic characteristics of 7 allotetraploid cotton species

Sample details, assembly statistics, methods and annotation details	***Gossypium hirsutum***	***Gossypium barbadense***	***Gossypium tomentosum***	***Gossypium mustelinum***	***Gossypium darwinii***	***Gossypium ekmanianum***	***Gossypium stephensii***
family	**Malvaceae**	**Malvaceae**	**Malvaceae**	**Malvaceae**	**Malvaceae**	**Malvaceae**	**Malvaceae**
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Synonym	*G. hirsutum var. punctatum, G. jamaicense, G. lanceolatum, G. mexicanum, G. morrillii, G. punctatum, G. purpurascens, G. religiosum, G. schottii, G. taitense, G. tridens, G. tricuspidatum, G. hirsutum var.* marie-galante and *G. hirsutum* var. *palmeri*	*G. peruvianum, G.vitifolium, G. acuminatum, G. barbadense var. acuminatum, G.barbadense var. brasiliense, G. brasiliense, G.guyanense var. braziliense, G. evertum*	*Gossypium hirsutum f. tomentosum* (Nutt. ex Seem.) Roberty, *Hibiscus tomentosus (Nutt. ex Seem.),*	*Gossypium hirsutum* subsp. *mustelinum* (Miers ex G.Watt) Roberty; *Gossypium caicoense Condorcet*	*Gossypium barbadense* var. *darwinii*	*Gossypium hirsutum* var. *ekmanianum* (Wittm.) Roberty, *Gossypium latifolium* var. *ekmanianum* (Wittm.) Roberty, and *Gossypium tricuspidatum* var. *ekmanianum* (Wittm.) Mauer	not applicable
Common name	American Tetraploid, American upland cotton, upland cotton	Pima cotton, American Pima cotton, Sea Island cotton, long-staple cotton, Egyptian cotton, Brazilian cotton	Hawaiian cotton	Brazilian cotton	Darwin's Cotton, Galapagos cotton	Ekman’s Cotton, Ekman’s *Gossypium,* and Ekman's Wild Cotton	Wake island cotton
Origin	Central America	South America	Hawaiian Islands	NE Brazi	Galapagos Islands	Dominican Republic	Wake Atoll, Pacific Ocean
Description	*G. hirsutum* is the most widly cultivated cotton species and a predominant source of natural plant fibers in the world. It is native to Central America but cultivated worldwide.	*G. barbadense* is a tropical cotton that grows to the size of a small tree. It is valued for its long, high-quality fiber and resistance to Verticillium wilt.	*G. tomentosum* has its own special agronomic characteristics, including strong fibers, hairy leaves and stems, nectarilessness on leaves or bracteoles, insect-pest resistance, and the most heat-tolerant species of Gossypium.	Small, thorny, deciduous, and xerophytic trees	Has finer fibers, shows resistance to drought, Fusarium and Verticillium wilt.	It is a shrub and grows primarily in the seasonally dry tropical biome. G. ekmanianum has small, white flowers with five petals. The seeds are small, round, and brown.	The new species morphological characteristics was well described by Gallagher, Grover et al. 2017.

Continuation of Table

1	2	3	4	5	6	7	8
Sequenced Cultivar	acc.TM-1	acc.3-79	isolate 7179.01,02,03	isolate1408120.09, 1408120.10,1408121 .01, 1408121.02, 1408121.03	(AD)5-032	AD6	AD7
Genome group	AD	AD	AD	AD	AD	AD	AD
Individual genome	$(AD)_1$	$(AD)_2$	$(AD)_3$	$(AD)_4$	$(AD)_5$	$(AD)_6$	$(AD)_7$
Haploid chromosomes number	26	26	26	26	26	26	26
Ploidy	allotetraploid	allotetraploid	allotetraploid	allotetraploid	allotetraploid	allotetraploid	allotetraploid
Individual chromosome	A_h01-A_h13; D_h01-D_h13	A_b01-A_b13; D_b01-D_b13	$A_{tt}01$-$A_{tt}13$; $D_{tt}01$-$D_{tt}13$	A_m01-A_m13; D_m01-D_m13	A_d01-A_d13; D_d01-D_d13	A_e01-A_e13; D_e01-D_e13	A_s01-A_s13; D_s01-D_s13
Development stage	young seedling	young seedling	young seedling	young seedling	young seedling	young seedling	young seedling
Tissue	leaf	leaf	leaf	leaf	leaf	leaf	leaf
Geographic location	USA	USA	USA	USA	USA	China	China
GenBank	CM017662.1 - CM017687.1	CM018202.1 - CM018227.1	CM017610.1 - CM017635.1	CM017636.1 - CM017661.1	CM017688.1 - CM017713.1	CM046631.1 - CM046656.1	CM045525.1 - CM045550.1
RefSeq	NC_053424.1- NC_053449.1	n/a	n/a	n/a	n/a	n/a	n/a
Submitter	HudsonAlpha Genome Sequencing Center	HudsonAlpha Genome Sequencing Center	HudsonAlpha Genome Sequencing Center	HudsonAlpha Genome Sequencing Center	HudsonAlpha Genome Sequencing Center	The institute of Cotton Research of Chinese Academy Agricultural Sciences	The institute of Cotton Research of Chinese Academy Agricultural Sciences
Date	Feb 25, 2021	Oct 8, 2019	Aug 27, 2019	Aug 27, 2020	Aug 27, 2019	Oct 6, 2022	Aug 17, 2022
Assembly type	haploid	haploid	haploid	haploid	haploid	haploid	haploid
Assembly level	Chromosome	Chromosome	Chromosome	Chromosome	Chromosome	Chromosome	Chromosome
Sequencing technology	Shotgun sequence, PacBio RSII	Shotgun sequence, PacBio RSII	Shotgun sequence, PacBio RSII	Shotgun sequence, PacBio RSII	Shotgun sequence, PacBio RSII	Shotgun sequence, PacBio Sequel	Shotgun sequence, PacBio Sequel
Assembly method	Mecat v. 1.0	Mecat v. 1.0	Mecat v. 1.1	Mecat v. 1.2	Mecat v. 1.1	FALCON v. 1	FALCON v. 1
Genome size	2.3 Gb	2.2 Gb	2.2 Gb	2.3 Gb	2.2 Gb	2.3 Gb	2.3 Gb
Total ungapped length	2.3 Gb	2.2 Gb	2.2 Gb	2.3 Gb	2.2 Gb	2.3 Gb	2.3 Gb
Genome Coverage	94.06x	90.1x	76.8x	94.1x	80.6x	106.0x	127.0x
Genes	75,376	74,561	78 281	74 660	78 303	74178	74,970
GC percent	34.5	34	34	34,5	34	34.5	34.5
chromosome A01	119,761,559	113 238 469	114 172 108	121 752 893	111 731 133	124 031 627	118 745 647
chromosome A02	108,141,443	99769429	105 044 580	109 867 862	102 188 074	109 822 344	108 788 785

ISSN 3080-6836 (Print) ISSN 3080-6844 (Online)

Continuation of Table

1	2	3	4	5	6	7	8
chromosome A03	113,693,209	105 981 974	106 988 191	114 197 739	108 106 663	114 352 146	110 887 122
chromosome A04	89,180,822	80 954 414	83 431 537	92 465 543	82 123 192	89 349 839	87 570 184
chromosome A05	111,098,753	102 458 744	104 743 809	113 805 608	105 499 638	124 669 549	113 168 133
chromosome A06	128,195,338	116 119 172	121 609 178	126 788 839	119 414 621	130 785 536	125 976 056
chromosome A07	98,902,531	93 754 744	96 764 941	99 644 675	95 097 007	98 644 386	98 532 636
chromosome A08	127,495,948	119 114 718	119 265 125	129 182 752	120 009 936	127 240 960	125 152 893
chromosome A09	85,335,976	77 140 527	79 219 835	82 652 960	79 385 658	84 756 691	83 529 980
chromosome A10	118,182,687	109 871 119	110 988 805	118 586 504	113 192 719	117 158 140	117 398 810
chromosome A11	124,181,751	114 694 469	120 496 074	126 800 230	115 039 066	132 145 079	122 787 908
chromosome A12	109,474,314	99 282 030	102 861 178	106 758 837	101 902 339	108 057 877	108 204 208
chromosome A13	111,646,624	108 235 369	110 351 505	115 677 886	110 366 452	115 830 965	114 611 390
chromosome D01	65,205,008	62 611 448	63 967 270	65 691 877	63 351 384	66 297 838	66 300 157
chromosome D02	72,186,496	66 877 699	69 897 195	71 596 151	68 431 162	73 472 059	72 813 539
chromosome D03	54,956,272	53 043 351	52 197 648	56 565 278	54 072 262	56 012 142	55 007 171
chromosome D04	58,229,188	54 342 187	54 488 862	57 383 358	54 473 739	62 932 355	58 287 303
chromosome D05	66,484,719	62 429 648	62 680 911	65 850 227	63 570 493	68 714 917	65 884 442
chromosome D06	66,684,206	62 820 931	63 218 559	66 874 005	63 279 081	68 485 142	66 657 206
chromosome D07	59,440,927	56 075 531	57 081 155	58 050 881	57 165 345	62 021 339	59 415 676
chromosome D08	69,427,147	65831471	67 500 665	69 705 406	66 245 736	71 546 767	69 992 981
chromosome D09	54,445,796	50 685 742	51 553 955	53 194 044	52 096 622	54 144 365	54 019 951
chromosome D10	68,089,194	65 066 651	65 609 545	67 709 701	65 927 692	70 211 476	69 146 489
chromosome D11	72,823,778	69 577 400	70 776 713	73 590 885	69 104 961	74 825 825	73 228 295
chromosome D12	65,099,798	59 787 526	60 769 583	63 954 031	61 226 500	64 434 879	62 814 374
chromosome D13	65,099,798	60 421 729	61 396 220	64 530 457	60 827 587	67 089 636	63 971 005
References	[12], [13], [16], [18]	[11], [18], [23]	[11], [18], [25]	[11], [18]	[11], [18]	[11], [35]	[11], [34], [35]

Author contributions

OA: data curation, formal analysis, software, and writing-original draft. MSA: supervision and editing. MSP: data resources, data curation, formal analysis, visualization. RMB: data curation, software. KB: data curation, software. TLA: analysis and writing. TD: data curation, writing — review and editing, conceptualization, formal analysis, investigation, supervision.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest.

Funding

This study was funded by the grant, titled "Creation of drought tolerance classification for Kazakhstan cotton-collection and identification of SNP markers associated with drought tolerance traits" (AP23489921).

Acknowledgments

We thank the Ministry of Science and Higher Education of the Republic of Kazakhstan for providing financial support.

References

1 Owusu, A.G. et al. (2023). Transcriptomic and metabolomic analyses reveal the potential mechanism of waterlogging resistance in cotton (Gossypium hirsutum L.). *Frontiers in Plant Science, 14,* 1088537.

2 Sun, K. et al. (2023). Transcriptome, proteome and functional characterization reveals salt stress tolerance mechanisms in upland cotton (Gossypium hirsutum L.). *Frontiers in Plant Science, 14,* 1092616.

3 Ma, Z. et al. (2021). High-quality genome assembly and resequencing of modern cotton cultivars provide resources for crop improvement. *Nature Genetics, 53, 9,* 1385–1391.

4 Statista. *statista.com.* Retrieved from https://www.statista.com/statistics/263055/cotton-production-worldwide-by-top-countries/

5 Economic Development Board. *just-style.com*. Retrieved from https://www.just-style.com/news/surge-in-2023-24-cotton-production-to-push-global-reserves-to-all-time-high/? cf-view

6 Holt, G., Barker, G., Baker, R., & Brashears, A. (2000). Characterization of cotton gin byproducts produced by various machinery groups used in the ginning operation. Transactions of the ASAE. *American Society of Agricultural Engineers, 43*, 1393–1400.

7 OECD-FAO AGRICULTURAL OUTLOOK. 2017–2026. © OECD/FAO 2017.

8 Makhmadjanov, S.P., Tokhetova, L.A., Daurenbek, N.M., Tagaev, A.M., & Kostakov, A.K. (2023). Cotton advanced lines assessment in the Southern Region of Kazakhstan. *SABRAO J. Breed. Genet., 55*(2), 279–290. http://doi.org/10.54910/sabrao2023.55.2.1

9 Kushanov, F.N. et al. (2022). Genetic analysis of mutagenesis that induces the photoperiod insensitivity of wild cotton Gossypium hirsutum subsp. purpurascens. *Plants, 11,* 22, 3012.

10 Grover, C.E. et al. (2015). Re-evaluating the phylogeny of allopolyploid Gossypium L. *Molecular Phylogenetics and Evolution, 92,* 45–52.

11 Wang, K., Wendel, J.F., & Hua, J. (2018). Designations for individual genomes and chromosomes in Gossypium. *Journal of Cotton Research, 1,* 1, 1–5.

12 Ning, W. et al. (2024). Origin and diversity of the wild cottons (Gossypium hirsutum) of Mound Key, Florida. *Scientific Reports, 14,* 1, 14046.

13 Li, F. et al. (2015). Genome sequence of cultivated Upland cotton (Gossypium hirsutum TM-1) provides insights into genome evolution. *Nature biotechnology, 33,* 5, 524–530.

14 Wang, M. et al. (2015). Long noncoding RNA s and their proposed functions in fibre development of cotton (Gossypium spp.). *New phytologist, 207,* 4, 1181–1197.

15 Hu, Y. et al. (2019). Gossypium barbadense and Gossypium hirsutum genomes provide insights into the origin and evolution of allotetraploid cotton. *Nature genetics, 51,* 4, 739–748.

16 Wang, M. et al. (2019). Reference genome sequences of two cultivated allotetraploid cottons, Gossypium hirsutum and Gossypium barbadense. *Nature genetics, 51,* 2, 224–229.

17 Yang, Z. et al. (2019). Extensive intraspecific gene order and gene structural variations in upland cotton cultivars. *Nature communications, 10,* 1, 2989.

18 Chen, Z.J. et al. (2020). Genomic diversifications of five Gossypium allopolyploid species and their impact on cotton improvement. *Nature genetics, 52,* 5, 525–533.

19 Huang, G. et al. (2020). Genome sequence of Gossypium herbaceum and genome updates of Gossypium arboreum and Gossypium hirsutum provide insights into cotton A-genome evolution. *Nature genetics, 52,* 5, 516–524.

20 Perkin, L.C. et al. (2021). Genome assembly of two nematode-resistant cotton lines (Gossypium hirsutum L.). *G3, 11,* 11, jkab276.

21 Cheng, Y. et al. (2024). Gossypium purpurascens genome provides insight into the origin and domestication of upland cotton. *Journal of Advanced Research, 56,* 15–29.

22 Sreedasyam, A. et al. (2024). Genome resources for three modern cotton lines guide future breeding efforts. *Nature Plants,* 1–13.

23 Shim, J., Mangat, P.K., & Angeles-Shim, R.B. (2018). Natural variation in wild Gossypium species as a tool to broaden the genetic base of cultivated cotton. *J. Plant Sci. Curr. Res., 2,* 005.

24 Yuan, D. et al. (2015). The genome sequence of Sea-Island cotton (Gossypium barbadense) provides insights into the allopolyploidization and development of superior spinnable fibres. *Scientific reports, 5*, 1, 17662.

25 Shen, C., Wang, N., Zhu, D., Wang, P., Wang, M., Wen, T., Le, Y., Wu, M., Yao, T., Zhang, X., & Lin, Z. (2021). Gossypium tomentosum genome and interspecific ultra-dense genetic maps reveal genomic structures, recombination landscape and flowering depression in cotton. *Genomics, 113,* 4, 1999–2009.

26 Liu, F., Zhou, Z.L., Wang, C.Y., Wang, Y.H., Cai, X.Y., Wang, X.X., Wang, K.B., & Zhang, Z.S. (2016). Collinearity analysis of allotetraploid Gossypium tomentosum and Gossypium darwinii. *Genetics and Molecular Research, 15,* 3.

27 Alves, M.F., Barroso, P.A., Ciampi, A.Y., Hoffmann, L.V., Azevedo, V.C., & Cavalcante, U. (2013). Diversity and genetic structure among subpopulations of Gossypium mustelinum (Malvaceae). *Genetics and Molecular Research, 12,* 1, 597–609.

28 Royal Botanic Gardens. *powo.science.kew.org.* Retrieved from https://powo.science.kew.org/taxon/urn: lsid: ipni.org: names:112728-2

29 Yang, Y., You, C., Wang, N., Wu, M., Le, Y., Wang, M., Zhang, X., Yu, Y., & Lin, Z. (2023). Gossypium mustelinum genome and an introgression population enrich interspecific genetics and breeding in cotton. *Theoretical and Applied Genetics, 136,* 6, 130.

30 Wang, B. et al. (2017). QTL analysis of cotton fiber length in advanced backcross populations derived from a cross between Gossypium hirsutum and G. mustelinum. *Theoretical and Applied Genetics, 130,* 1297–1308.

31 Selina Wamucii. *selinawamucii.com.* Retrieved from https://www.selinawamucii.com/plants/malvaceae/gossypium-ekmanianum/#description

32 Stephens, S. (1966). The potentiality for long-range oceanic dispersal of cotton seeds. *The American Naturalist, 100*, 912, 199–210.

33 Fryxell, P.A. (1992). A revised taxonomic interpretation of Gossypium L. (Malvaceae). *Rheedea, 2,* 2, 108–165.

34 Gallagher, J., Grover, C., Rex, K., Moran, M., & Wendel, J. (2017). A new species of cotton from Wake Atoll, Gossypium stephensii (Malvaceae). *Systematic Botany, 42*, 115–123.

35 Peng, R., Xu, Y., Tian, S., Unver, T., Liu, Z., Zhou, Z., Cai, X., Wang, K., Wei, Y., Liu, Y., Wang, H., Hu, G., Zhang, Z., Grover, C.E., Hou, Y., Wang, Y., Li, P., Wang, T., Lu, Q., Wang, Y., Conover, J.L., Ghazal, H., Wang, Q., Zhang, B., Van Montagu, M., Van de Peer, Y., Wendel, J.F., & Liu, F. (2022). Evolutionary divergence of duplicated genomes in newly described allotetraploid cottons. *Proceedings of the National Academy of Sciences of the United States of America, 119*, 39, e2208496119.

36 Wang, W., Li, Y., Le, M., Tian, L., Sun, X., Liu, R., Guo, X., Wu, Y., Li, Y., Zhao, J., Liu, D., & Zhang, Z. (2024). QTL Mapping of Fiber- and Seed-Related Traits in Chromosome Segment Substitution Lines Derived from Gossypium hirsutum × Gossypium darwinii. *International Journal of Molecular Sciences, 25*, 17, 9639.

37 Xu, C., Ilyas, M.K., Magwanga, R.O., Lu, H., Khan, M.K.R., Zhou, Z., Li, Y., Kuang, Z., Javaid, A., Ibrar, D., Ghafoor, A., Wang, K., Liu, F., & Chen, H. (2022). Transcriptomics for drought stress mediated by biological processes in relation to key regulated pathways in *Gossypium darwinii. Molecular Biology Reports, 49,* 12, 11341–11350.

38 Chen, H., Khan, M.K., Zhou, Z., Wang, X., Cai, X., Ilyas, M.K., Wang, C., Wang, Y., Li, Y., Liu, F., & Wang, K. (2015). "A high-density SSR genetic map constructed from a F2 population of Gossypium hirsutum and Gossypium darwinii". *Gene, 574*(2), 273–286.

А. Өркен, Ш. Манабаева, С. Махмаджанов, М. Рамазанова, Б. Қали,
Л. Тохетова, Д. Түсіпқан

Мақта (*Gossypium L.*) өндірісі және агрономиялық сипаттамаларын жақсарту үшін секвенирлеу технологиясының маңызы

Gossypium L. — дақылдардың әртүрлілігімен және экономикалық құндылығымен танымал ең ірі тұқымдардың бірі, ал аллотетраплоидты мақта түрлері өсімдіктердің полиплоидиясын, филогенезін және селекциясын зерттеудің құнды көзі және модельдік жүйесі. Бұл шолуда, біріншіден, әлемдегі

мақта (*Gossypium L.*) өндірісі мен қолданылуы туралы ақпарат қамтылған. Екіншіден, Қазақстанда мақта ауыл шаруашылығы және мақта өндірісі туралы толық ақпарат ұсынылған. Үшіншіден, *Gossypium L.* филогенезі туралы қысқаша ақпарат берілген. Төртіншіден, *G. hirsutum* (AD)1, *G. barbadense* (AD)2, *G. tomentosum* (AD)3, *G. mustelinum* (AD)4, *G. darwinii* (AD)5, *G. ekmanianum* (AD)6 және *G. stephensii* (AD)7 сияқты жеті аллотетраплоидты мақта түрлерінің толық геномының морфологиялық сипаттамалары мен зерттеулеріне қысқаша шолу берілген. Бұл шолу мақтаның агрономиялық және молекулалық зерттеулері үшін құнды ақпаратты ұсынады.

Кілт сөздер: мақта өндірісі (Gossypium L.), филогенез, аллотетраплоидты түрлер, геномның толық тізбегі.

А. Өркен, Ш. Манабаева, С. Махмаджанов, М. Рамазанова, Б. Қали,
Л. Тохетова, Д. Түсіпқан

Производство хлопчатника (*Gossypium L.*) и важность технологий секвенирования для улучшения агрономических характеристик хлопка

Gossypium L. — один из крупнейших родов, известных своим разнообразием и экономической ценностью среди сельскохозяйственных культур, в то время как аллотетраплоидные виды хлопчатника являются ценным источником и модельной системой для изучения полиплоидии растений, филогении и селекции. Этот обзор включает, во-первых, информацию о производстве и использовании хлопчатника (*Gossypium*) в мире. Во-вторых, подробные сведения о сельском хозяйстве и производстве хлопчатника в Казахстане. В-третьих, приведена сводка о филогении *Gossypium L.* В-четвертых, представлен краткий обзор морфологических характеристик и исследований полного генома семи аллотетраплоидных видов хлопчатника, включая *G. hirsutum* (AD)1, *G. barbadense* (AD)2, *G. tomentosum* (AD)3, *G. mustelinum* (AD)4, *G. darwinii* (AD)5, *G. ekmanianum* (AD)6 и *G. stephensii* (AD)7. Данный обзор представляет собой ценный источник информации для будущих агрономических и молекулярных исследований хлопчатника.

Ключевые слова: Производство хлопчатника (*Gossypium* L.), филогенез, аллотетраплоидные виды, полная последовательность генома.

Information about the authors

Orken Aisulu — Master student in Biology Sciences, Laboratory Assistant of the Plant Genetic Engineering Laboratory, National Center for Biotechnology, Astana, Kazakhstan; e-mail: *orkena23@gmail.com*

Manabayeva Shuga Askarovna — PhD in Biology Sciences, Head of the Plant Genetic Engineering Laboratory, National Center for Biotechnology, Astana, Kazakhstan; e-mail: *manabayeva@biocenter.kz*

Makhmadjanov Sabir Partovich — PhD in Agricultural Sciences, Head of the Department of Transfer and Adaptation of Crop Varieties, Agricultural Experimental Station of Cotton and Melon Growing; e-mail: *max_s1969@mail.ru*

Ramazanova Malika Baglanovna — Master in Natural Sciences, Junior researcher of the Plant Genetic Engineering Laboratory, National Center for Biotechnology, Astana, Kazakhstan; e-mail: *malikaramazan.7@gmail.com*

Kali Balnur — Master in Natural Sciences, Researcher of the Plant Genetic Engineering Laboratory, National Center for Biotechnology, Astana, Kazakhstan; e-mail: *kali@biocenter.kz*

Tokhetova Laura Anuarovna — Doctor of Agricultural Sciences, Principal researcher, Agricultural Experimental Station of Cotton and Melon Growing; e-mail: *lauramarat_777@mail.ru*

Tussipkan Dilnur — PhD in Crop genetics and Breeding, Leading researcher of the Plant Genetic Engineering Laboratory, National Center for Biotechnology, Astana, Kazakhstan; e-mail: *tdilnur@mail.ru*

 ISSN 3080-6836 (Print) ISSN 3080-6844 (Online)